Toward Optimality in Discrete Morse Theory

Thomas Lewiner, Hélio Lopes, and Geovan Tavares

CONTENTS

2000 AMS Subject Classification: Primary 58E05, 68U05; Secondary 57M15, 57M27, 57Q10

Keywords: Morse theory, Forman theory, computational topology, computational geometry, solid modeling, discrete mathematics

Morse theory is a fundamental tool for investigating the topology of smooth manifolds. This tool has been extended to discrete structures by Forman, which allows combinatorial analysis and direct computation. This theory relies on discrete gradient vector fields, whose critical elements describe the topology of the structure. The purpose of this work is to construct optimal discrete gradient vector fields, where optimality means having the minimum number of critical elements. The problem is equivalently stated in terms of maximal hyperforests of hypergraphs. Deduced from this theoretical result, a algorithm constructing almost optimal discrete gradient fields is provided. The optimal parts of the algorithm are proved, and the part of exponential complexity is replaced by heuristics. Although reaching optimality is MAX-SNP hard, the experiments on odd topological models are almost always optimal.

1. INTRODUCTION

1.1 Morse Theory

Morse theory [Milnor 63] is a fundamental tool for investigating the topology of smooth manifolds. Morse proved that the topology of a manifold is very closely related to the critical points of a real smooth map defined on it. The simplest example of this relationship is the fact that if the manifold is compact, then any continuous function defined on it must have a maximum and a minimum. Morse theory provides a significant refinement of this observation.

1.2 Forman's Discrete Morse Theory

The recent insights in Morse theory by Forman [Forman 95, Forman 98] extended several aspects of this fundamental tool to discrete structures. Its combinatorial aspect allows computation completely independent of a geometric realization: The algorithms we designed do not require any coordinate or floating-point calculation, and geometrical constraints can be applied independently. Forman proved several results and provided many applications of his theory [Forman 00, Forman 01]. Once a Morse function has been defined on a CW-complex, then

information about its topology can be partly deduced from its critical cells (i.e., where the discrete gradient vanishes).

1.3 Optimality

Similarly to the differential case, Forman proved that the topology of a CW-complex can be partly read out of the critical cells of a discrete gradient vector field defined on it. The topological information will be concise if the discrete gradient vector field has few critical cells. Hence, we will say that a gradient vector field is *optimal* if it has the minimum possible number of critical cells.

1.4 Results

The goal of this work is to build an optimal gradient vector field. To do so, we develop in Section 4 a hypergraph representation of a discrete gradient vector field. We introduce the notion of hyperforest (Section 4.3) and prove the equivalence between discrete gradient vector fields and hyperforests in Theorem 4.7. We stated the equivalent of a critical cell for a hyperforest in Proposition 4.9. We finally prove that the minimum number of critical cells is a topological invariant for 3-manifolds in Theorem 4.13.

We provide in Section 5 an algorithm to build a discrete gradient vector field on general cell complexes of arbitrary dimension. This algorithm is worst-case quadratic in execution time. It is not guaranteed to be optimal, but it gives optimal results in most of the cases (see results in Section 6.3). For the particular case of 2-manifolds, the algorithm is proven to be optimal, although the general problem is MAX-SNP hard[1] [Lewiner 02].

Sections 2 and 3 recall some basic definitions of graph theory and discrete Morse theory. The theoretical base of our work is detailed in Section 4. Our construction and its optimality are discussed in Section 5. As the problem is NP hard, we used some heuristics which are compared to their experimental results in Section 6.

2. DISCRETE STRUCTURES

2.1 Cell Complexes

A cell complex is, roughly speaking, a generalization of the structures used to represent solid models: It is a consistent collection of cells (vertices, edges, faces ...). In particular, triangulations of topological spaces or three-dimensional meshes are cell complexes (see Figure 1).

[1] A MAX-SNP hard problem is an NP hard problem for which any polynomial algorithm can lead arbitrary far from the optimum.

FIGURE 1. A triangulated torus.

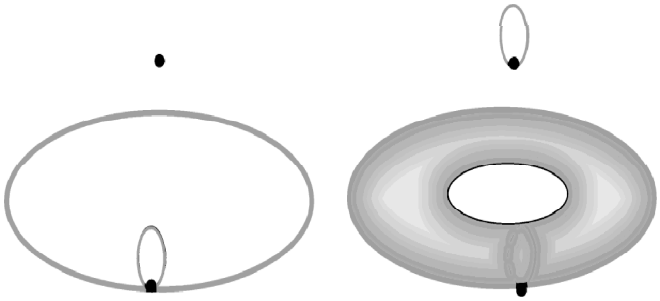


FIGURE 2. A construction of a torus with four cells.

Figure 2 gives a minimal construction of a torus cell complex. A complete introduction to cell complexes can be found in [Lundell 69].

Definition 2.1. (Cell.) A *cell* $\alpha^{(p)}$ of dimension p is a set homeomorphic to the open p–ball $D^p = \{x \in \mathbb{R}^p : \|x\| < 1\}$.

When the dimension p of the cell is obvious, we will simply denote α instead of $\alpha^{(p)}$. Those cells are attached together to form a cell complex, in the following way:

Definition 2.2. (Attaching a cell.) Given a p–cell α, a set $X^{(p-1)}$ of (p-1)–cells and a map $\phi_\alpha : \partial D^p \to X^{(p-1)}$, the attachment of α to $X^{(p-1)}$ via ϕ_α is the quotient space of the disjoint union $X^{(p-1)} \sqcup D^p$ under the identification $x \equiv \phi_\alpha(x)$.

Definition 2.3. (CW-complex.) A *CW-complex* K is built by starting off with a discrete collection of 0-cells (vertices) called K^0, then attaching 1-cells (edges) to K^0 along their boundaries, obtaining K^1, then attaching 2-cells (faces) to K^1 along their boundaries, writing K^2 for the new space, and so on, giving spaces K^n for every n.

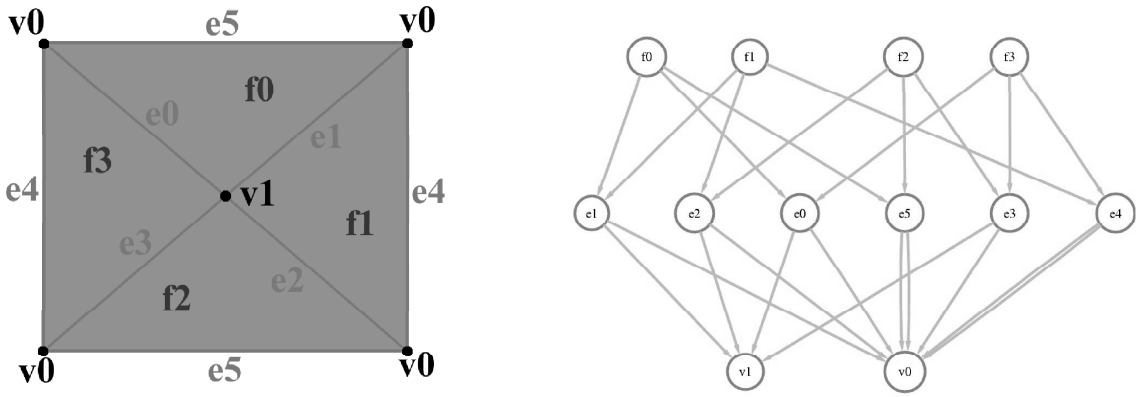


FIGURE 3. The Hasse diagram of a non-PL torus: The links $e5 \to v0$ and $e4 \to v0$ are double.

A CW-complex will be said to be finite when it is built out of a finite number of cells. In this work, we will consider only finite (and thus regular) CW-complexes, in order to compute on them.

A p-cell $\alpha^{(p)}$ is a face of a q-cell $\beta^{(q)}$ $(p < q)$ if $\alpha \subset closure(\beta)$. If $q = p + 1$, we will use the notation $\alpha^{(p)} \prec \beta^{(q)}$, and say that α and β are *incident*. The multiplicity of this incidence is the number of connected components of $\phi_\beta^{-1}(\alpha)$.

In a sense, a cell complex is a generalization of a graph, as a graph can be seen as a cell complex of dimension 1. Nevertheless, we can also represent a cell complex by a multigraph[2], called its *Hasse diagram*.

Definition 2.4. (Hasse diagram.) The *Hasse diagram* of a cell complex K is the oriented multigraph H, where:

- Each node of H represents a cell of K.
- The links of H join nodes representing incident cells of K. Multiple incidences are represented by multiple links. The source of each link is the one of highest dimension.

The Hasse diagram is usually drawn with the nodes ranked by their dimension. In Figure 3, the faces (2-cells) are aligned on top rank, the edges (1-cells) on the middle one, and the vertices (0-cells) on the bottom rank. A link between two nodes symbolizes that the corresponding cells are incident.

2.2 Hypergraphs

In the dual graph of a cell complex that is not a manifold, links that join more than two nodes may appear. This would not fit in the definition of a simple graph, but in the following one:

Definition 2.5. (Hypergraph.) A *hypergraph* is a pair (N, L), where L is a family of families of N. The elements of L are called *hyperlinks*.

The elements of L are families, which means that a hyperlink can be incident more than once to a node. They can also be empty. We will classify nonempty hyperlinks into the *regular hyperlinks* (or shortly *link*), which join two distinct nodes as in simple graphs, the *loops*, which are incident to only one node, and the *nonregular hyperlinks*, which either join three or more nodes or are multiply incident to one node. We can extract the simple graph part of a hypergraph by considering its *regular components*:

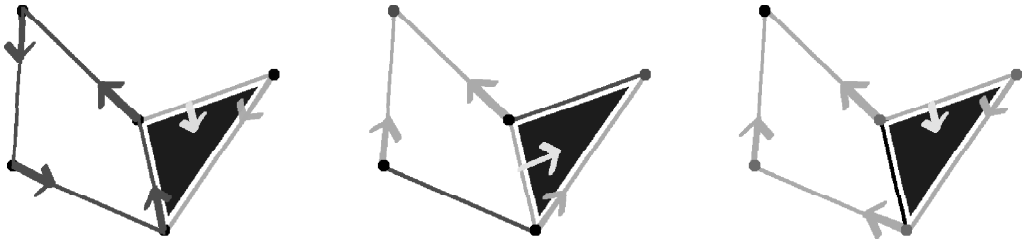
Definition 2.6. (Regular components.) The *regular components* of a hypergraph (N, L) are the connected components of the simple graph (N, R), where R is the set of the regular hyperlinks of (N, L).

Multiple incidences (i.e., duplicated edges) are reduced inside the regular components. We will give a hypergraph a simple orientation by distinguishing one node of each hyperlink as its *source*.

A hypergraph can be represented by a bipartite graph [Berge 70]. This gives a simple, but expensive, representation of hypergraphs:

Definition 2.7. (Bipartite graph of a hypergraph.) The *bipartite graph* $\mathcal{B}(H)$ of a hypergraph $H = (N, L)$ is the simple graph whose two classes of nodes are $\{\mathcal{B}(n), n \in N\}$ and $\{\mathcal{B}(l), l \in L\}$, representing the nodes and the links of H.

[2] A multigraph admits multiple links.



(a) combinatorial vector field (b) valid discrete gradient field (c) optimal discrete gradient field

FIGURE 4. Example of a combinatorial vector fields: (a) not a gradient (the closed $\mathcal{V}$–path is highlighted); (b) and (c) are gradient vector fields; (c) is optimal.

For every hyperlink $l = (n_1, n_2, \ldots, n_k)$ of H, there are k links of $\mathcal{B}(H)$ joining $\mathcal{B}(l)$ to $\mathcal{B}(n_i)$ (with $i = 1 \ldots k$).

When H is oriented, $\mathcal{B}(H)$ will be oriented the following way:

- If a node n of H is the source of a hyperlink l, then $\mathcal{B}(l)$ will be the source of the link of $\mathcal{B}(H)$ joining $\mathcal{B}(n)$ to $\mathcal{B}(l)$.
- If a node n of H is not the source of an incident hyperlink l, then $\mathcal{B}(n)$ will be the source of the link of $\mathcal{B}(H)$ joining $\mathcal{B}(n)$ to $\mathcal{B}(l)$.

The operation of taking the bipartite graph of a hypergraph can be reversed. Depending on which class of nodes becomes the links of the hypergraph, we can obtain a hypergraph or its dual. The bipartite graph is not supposed to give a consistent orientation in the general case. Therefore, the hypergraph representing a bipartite graph will not always be oriented.

Definition 2.8. (Hypergraphs of a bipartite graph.) A bipartite graph B admits two representations by hypergraphs: $\mathcal{B}^{-1}(B)$ and its dual $\mathcal{D}\left(\mathcal{B}^{-1}(B)\right)$. The nodes of $\mathcal{B}^{-1}(B)$ represent the nodes of one class of B, and the hyperlinks of $\mathcal{B}^{-1}(B)$ represent the nodes of the other class. For every node l of the second class, there is a hyperlink of $\mathcal{B}^{-1}(B)$ joining all the nodes adjacent to l.

3. FORMAN'S DISCRETE MORSE THEORY

3.1 Discrete Gradient Vector Field

Definition 3.1. (Combinatorial vector field.) A *combinatorial vector field* $\mathcal{V}$ defined on a cell complex K is a collection of disjoint pairs of incident cells $\{\alpha^{(p)} \prec \beta^{(p+1)}\}$.

We will represent a combinatorial vector field by an arrow from the cell of lower dimension to its paired cell of higher dimension (see Figure 4).

Definition 3.2. ($\mathcal{V}$-path.) A *$\mathcal{V}$-path* is an alternating sequence of cells $\alpha_0^{(p)}, \beta_0^{(p+1)}, \ldots, \alpha_r^{(p)}, \beta_r^{(p+1)}, \alpha_{r+1}^{(p)}$ satisfying: $\{\alpha_i^{(p)} \prec \beta_i^{(p+1)}\} \in \mathcal{V}$, $\beta_i^{(p+1)} \succ \alpha_{i+1}^{(p)}$ and $\alpha_{i+1}^{(p)} \neq \alpha_i^{(p)}$.

A $\mathcal{V}$-path is nontrivial and closed if $r \geq 1$ and $\alpha_{r+1} = \alpha_0$. For example, Figure 4(a) shows in black the closed $\mathcal{V}$-path of a combinatorial vector field.

Definition 3.3. (Discrete gradient vector field.) A *discrete gradient vector field* is a combinatorial vector field with no nontrivial closed $\mathcal{V}$-path.

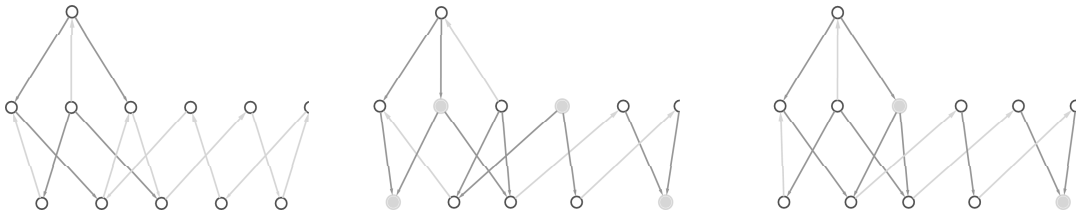
3.2 Critical Cells

Morse proved that the topology of a manifold is related to the critical elements of a smooth function defined on it. Forman gave a similar result, with the following definition for the critical cells:

Definition 3.4. (Critical cells.) A cell α is a *critical cell* of $\mathcal{V}$ if it is not paired with any other cell in $\mathcal{V}$.

The example of Figure 4 (a) is not a discrete gradient vector field as it contains a closed $\mathcal{V}$-path. In Figures 4(b) and 4(c), the critical cells of the discrete gradient vector field are drawn in black.

The number of critical cells is not a topological invariant of the cell complex, as it depends on the discrete gradient vector field considered. For example, with an empty discrete vector field (i.e., no cells are paired), every cell is critical, which would be the maximal number



(a) combinatorial vector field (b) valid discrete gradient field (c) optimal discrete gradient field

FIGURE 5. Hasse diagrams of the examples of Figure 4: (a) not a gradient (the circuit of its closed $\mathcal{V}$-path is highlighted); (b) and (c) are gradient vector fields; (c) is optimal.

of critical cells. In this work, we are more concerned with minimizing this number, as it would give a more concise description of the topology.

3.3 Hasse Diagram of Vector Fields

A combinatorial vector field is a partial matching in the Hasse diagram: Each pair of $\mathcal{V}$ corresponds to matched nodes in the Hasse diagram.

We will represent such a matching by inverting the orientation of the link between each pair in $\mathcal{V}$: For each $\{\alpha^{(p)} \prec \beta^{(p+1)}\} \in \mathcal{V}$ the source of the link joining $\alpha^{(p)}$ and $\beta^{(p+1)}$ in the diagram will be $\alpha^{(p)}$. For example, Figure 5 shows the Hasse diagrams of the discrete gradient vector fields of Figure 4. With this modified orientation, a closed $\mathcal{V}$-path is just an oriented circuit in the Hasse diagram (see Figure 5(a)).

A discrete gradient vector field contains no closed $\mathcal{V}$-path, and thus will be an *acyclic partial matching.*

4. STRUCTURE OF A DISCRETE GRADIENT VECTOR FIELD

A discrete gradient vector field has been defined as an acyclic partial matching in the Hasse diagram (see Section 3.3). This involves two problems: creating a matching, and removing cycles. Those two problems are separately well understood (see [Lovász 86] for matching theory, and [Hopcroft 73] for graph algorithms). However, when combined, they create NP hard problems [Lewiner 02]. In this section, we will give another point of view on discrete Morse theory in terms of the simplest (linear instead of quadratic) of those two problems: creating forests. We will prove our combined problem can be seen as a *hyperforest* creation problem.

4.1 Layers of the Hasse Diagram

In an n-combinatorial manifold, a (n-1)-cell is incident to either 1 or 2 n-cells [Lundell 69]. So the dual layer $n/(n\text{-}1)$ of the Hasse diagram will be represented by a pseudograph[3], called the dual pseudograph. This pseudograph can be seen as the hypergraph representation of the *dual layer* $n/(n\text{-}1)$ of the Hasse diagram.

Definition 4.1. (Layer of the Hasse diagram.) For two consecutive dimensions p and q of a cell complex K ($|p-q|=1$), the *layer* $\mathcal{L}_{p/q}$ of rank p/q of its Hasse diagram is an oriented simple bipartite graph. Its classes of nodes are the p- and q-cells of K. Its links join nodes representing incident p- and q-cells of K.

This definition distinguishes $\mathcal{L}_{p/q}$ from $\mathcal{L}_{q/p}$, which are dual following Definition 2.8. The orientation of those layers is the same as the one of the original Hasse diagram. For example, Figure 6 shows a double cube, its Hasse diagram, and the hypergraph of the bipartite graph $\mathcal{L}_{2/1}$ (see Definition 2.8), i.e., the dual graph of this double cube.

4.2 Reduced Layer of a Combinatorial Vector Field

Considering successive layers of the Hasse diagram is redundant: Each p-cell of K appears in general in four layers ($\mathcal{L}_{p/(p+1)}$, $\mathcal{L}_{p/(p-1)}$, $\mathcal{L}_{(p+1)/p}$, $\mathcal{L}_{(p-1)/p}$). When the Hasse diagram is oriented by a discrete gradient vector field (see Section 3.3), a matching belongs to only two of them. The following reduction allows such partition:

Definition 4.2. (Reduced layers of a combinatorial vector field.) Let K be a cell complex, $\mathcal{V}$ a combinatorial vector field defined on it, and $\mathcal{L}_{p/q}$ the layer p/q of the Hasse diagram oriented by $\mathcal{V}$, with $|p-q|=1$. The *reduced layer* $L_{p/q}$ is an oriented bipartite graph obtained by removing from $\mathcal{L}_{p/q}$:

- the p-cells of K paired with a q'-cell of K in $\mathcal{V}$, $q' \neq q$.
- the q-cells of K unpaired or paired with a p'-cell in $\mathcal{V}$, $p' \neq p$.

[3] A pseudograph admits loops and multiple links.

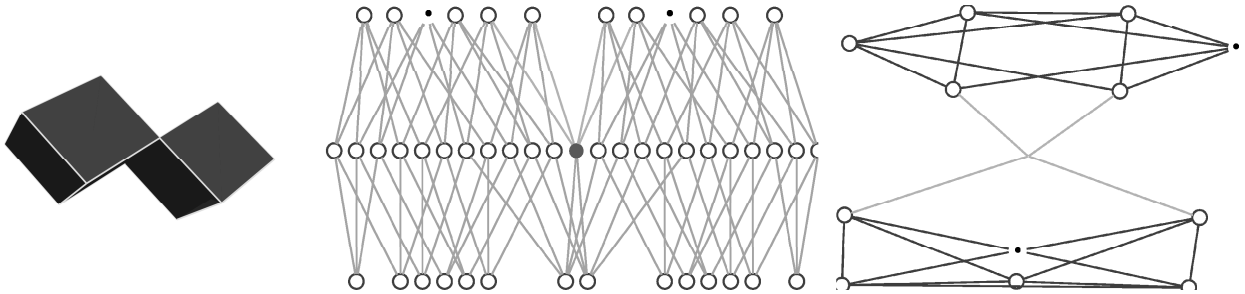
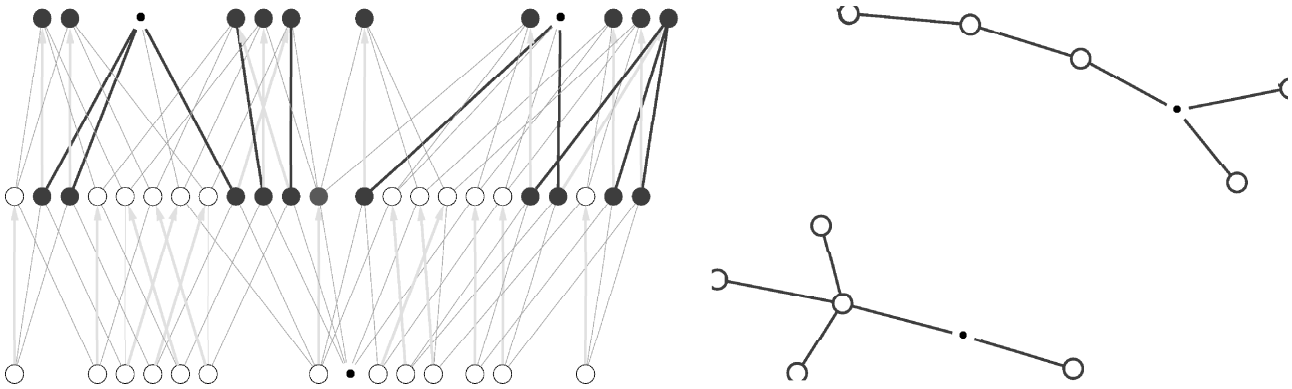


FIGURE 6. The layer of a double cube made with 12 squared faces, 23 edges, and 14 vertices. The hypergraph of the layer 2/1 of its Hasse diagram has therefore 12 nodes and 23 links.



(a) The reduced layer 2/1 of the double cube (dark nodes). (b) The hypergraph of the reduced layer 2/1.

FIGURE 7. Given the gradient field (dark lines) (a), its reduced layer is a forest (b).

Notice that any $\mathcal{V}$-path is entirely represented in one of the reduced layers. For example, Figure 7 shows in the dark dots the edges in the Hasse diagram of Figure 6 that belong to the reduced layer 2/1. The corresponding hypergraph is a forest (see Figure 7).

4.3 Hyperforests

A forest is a graph with no circuit. We now give a natural extension of forests for hypergraphs:

Definition 4.3. (Oriented hypercircuit.) An *oriented hypercircuit* in a hypergraph is a sequence of distinct nodes $n_0, n_1, \ldots, n_{r+1}$ such that $n_{r+1} = n_0$ and for all $0 \leq i \leq r$, n_i is the source of a hyperlink leading to n_{i+1}.

Definition 4.4. (Hyperforest.) We will say that a simply oriented hypergraph is a *hyperforest* if each node is the source of at most one hyperlink, and if it does not contain any hypercircuit.

In Figure 8, the oriented links belong to a discrete gradient vector field of the nonmanifold object drawn. The loops and nonregular hyperlink end the regular components.

Proposition 4.5. *Let HF be a hyperforest, and R one of its regular components.*

(i) The regular components of a HF are simple trees.

(ii) There is at most one node in R which is the source of either a loop or nonregular hyperlink.

Proof: (i). Suppose R had a (simple) circuit $n_0, n_1, \ldots, n_{r+1} = n_0$. There are $(r+1)$ nodes and $(r+1)$ regular links in this circuit. As a node cannot be the source of two links, each node is the source of exactly one link of the circuit.

Suppose, without loss of generality, that n_0 is the source of the link $\{n_0, n_1\}$. n_1 is incident to two links of the circuit: $\{n_0, n_1\}$ and $\{n_1, n_2\}$. As it is not the source of the first one, it is the source of $\{n_1, n_2\}$. Continuing those deductions, we prove that all the links of the circuit are oriented in such a way to form an oriented hypercircuit.

Since HF is a hyperforest, this leads to a contradiction. Therefore, R is a simple tree.

(ii). Let k be the number of nodes of R. As R is a tree, it has $(k-1)$ (regular) links. The sources of those links

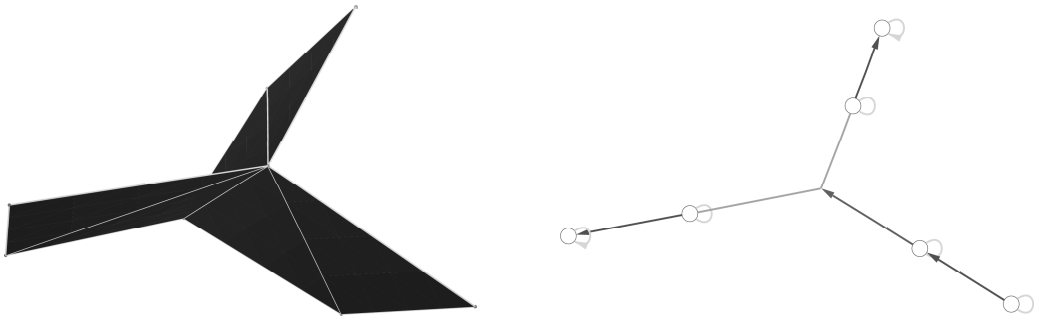
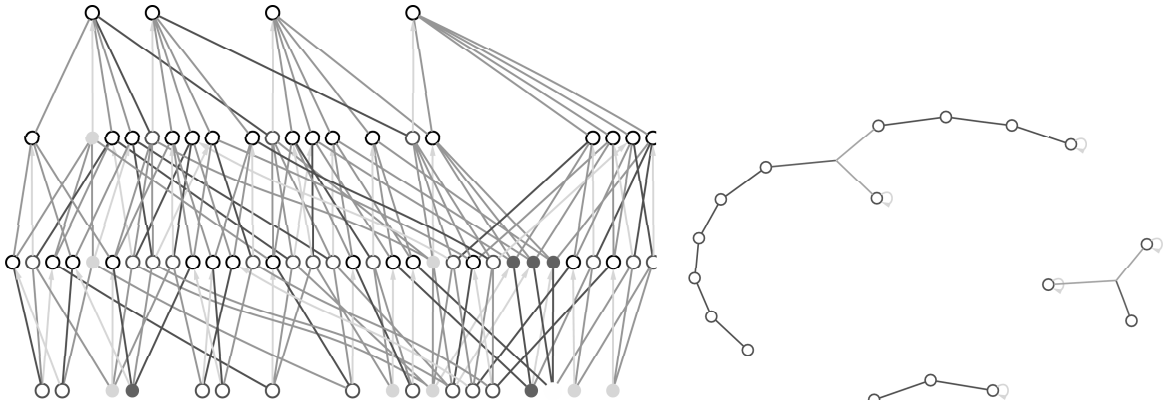


FIGURE 8. The nonregular hyperlinks appear while processing nonmanifold surfaces. Loops appear while processing surfaces with boundary.



(a) (b)

FIGURE 9. (a) The Hasse diagram of a discrete gradient vector field on a 4 cubes solid model, with a discrete gradient vector field defined on it (darker lines); (b) The 0/1 hyperforest of its reduced layer.

are nodes of R, as those links are regular (see Definition 2.6). Therefore, among those k nodes, there are $(k-1)$ nodes that are the source of regular hyperlinks. So there is at most $k-(k-1)=1$ node in R which is the source of either a loop or a nonregular hyperlink. □

4.4 Discrete Gradient Vector Field and Hyperforests

We defined a discrete gradient vector field as an acyclic partial matching in the Hasse diagram (see Section 3.3), and a hyperforest as a hypergraph without hypercircuit. Therefore, each layer of the Hasse diagram oriented by a discrete gradient vector field will have no hypercircuit. Hence, their representations by hypergraphs will be hyperforests.

Definition 4.6. (Hypergraphs of a combinatorial vector field.) Let K be a cell complex, $\mathcal{V}$ a combinatorial vector field defined on it, and $L_{p/q}$ the reduced layer p/q of $\mathcal{V}$ ($|p-q|=1$). The p/q–hypergraph of $\mathcal{V}$, $H_{p/q}$, is the hypergraph representation of $L_{p/q}$: $H_{p/q}=\mathcal{B}^{-1}(L_{p/q})$. $H_{p/q}$ is oriented as follow: The source of a hyperlink of $H_{p/q}$ is the node representing its paired cell in $\mathcal{V}$.

For example, Figure 9 shows the hyperforest corresponding to the reduced layer 1/0 of the Hasse diagram.

Theorem 4.7. *Let $\mathcal{V}$ be a combinatorial vector field. $\mathcal{V}$ is a discrete gradient vector field on an n-cell complex K if and only if the 0/1, 1/2, ... (n-1)/n hypergraphs of $\mathcal{V}$ are hyperforests.*

As the dual of a hyperforest is a hyperforest, the theorem is valid for any sequence obtained by replacing any of the p/q-hypergraphs by a q/p-hypergraph of $\mathcal{V}$.

Proof: The orientation of $L_{p/q}$ ensures the first condition of Definition 4.4. As any $\mathcal{V}$-path is entirely represented in one of the reduced layers, we just need to prove that a closed $\mathcal{V}$-path is a hypercircuit in one of the hypergraphs.

Let $n_0, n_1, \ldots, n_{r+1} = n_0$ be an oriented hypercircuit in $H_{p/q}$. From Definition 4.3, n_i is the source of a hyperlink l_i incident to n_{i+1}. This hyperlink l_i represents a q-cell β_i of K, and n_i represents a p-cell α_i. As n_i is the source of l_i, we know from the orientation of Definition 4.6 that α_i and β_i are incident and form a pair in $\mathcal{V}$. So $\alpha_0, \beta_0, \ldots, \alpha_r, \beta_r, \alpha_{r+1}$ is a $\mathcal{V}$-path. As $n_{r+1} = n_0$ and $r \geq 1$, this is a closed $\mathcal{V}$-path.

This argument can be reversed to prove that a closed $\mathcal{V}$-path is hypercircuit in one of the p/q-hypergraphs $H_{p/q}$. □

We will now define the analog of critical cells for hyperforests. This will be the foundation of the algorithm of Section 5. A critical cell of a discrete gradient vector field corresponds to one of the regular components of one of its hyperforests.

Definition 4.8. (Critical component.) A regular component of a hyperforest will be called *critical* if none of its nodes is the source of either a loop or a nonregular hyperlink.

Proposition 4.9. *Let $H_{p/q}$ be the p/q-hyperforest of a discrete gradient vector field $\mathcal{V}$. The number of critical components of $H_{p/q}$ is exactly the number of critical p-cells of $\mathcal{V}$.*

Proof: Every possible critical p-cell is represented $H_{p/q}$ and its corresponding reduced layer $L_{p/q}$. The isolated nodes of $L_{p/q}$ are not matched with any cell of K, and remain isolated nodes in $H_{p/q}$. Those nodes are critical components, according to Definition 4.8.

We know from Proposition 4.5 that each regular component R is a simple tree. In such a tree with k nodes, there are $(k-1)$ (regular) links. All links are oriented, so among those k nodes, $(k-1)$ are the sources of links of R, and therefore, those are not critical. If R is not a critical component, there is exactly one node of R which is the source of either a loop or a nonregular hyperlink, i.e., it is not critical.

If R is a critical component, this node is neither the source of a loop nor of a nonregular hyperlink. From Definition 2.6 of a regular component, this node is not incident to any regular hyperlink not in R. All those links of R are already paired with other nodes. So this node is unpaired in $L_{p/q}$. From Definition 4.2, it cannot be paired with a cell outside $L_{p/q}$. Therefore, it is an unpaired node, i.e., a critical cell. □

4.5 Optimality of Hyperforests

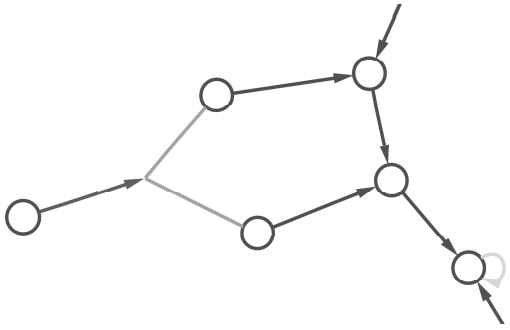


FIGURE 10. Detail of a hyperlink insertion in the dual hyperforest appearing with a solid torus model.

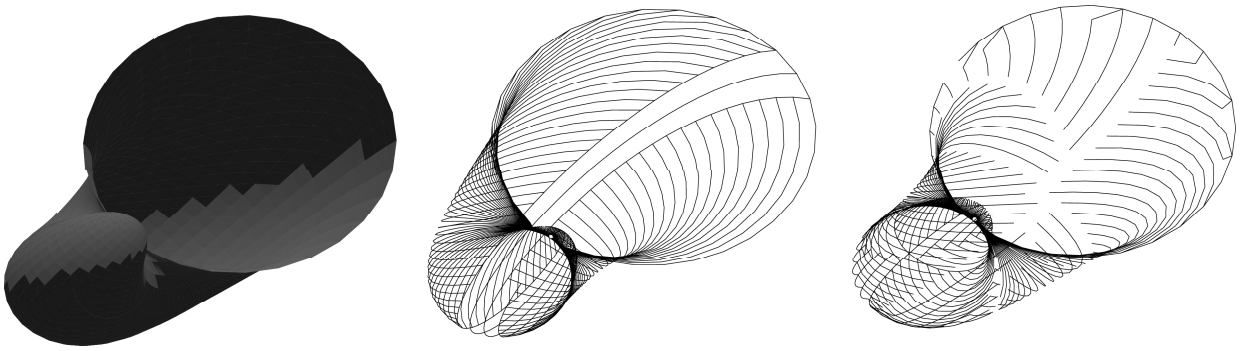
An optimal discrete gradient vector field will have the minimal possible number of critical cells, or equivalently, the minimum possible sum of critical components in each hyperforest $H_{p/q}$. There are as many noncritical elements in a hyperforest as its number of hyperlinks (noncritical elements are paired with an incident hyperlink). Therefore, an optimal discrete gradient vector field has the maximum total number of hyperlinks in all of its hyperforests. For example, adding the hyperlink on the left side of Figure 10 allows us to pair it with the node on the left. Thus, there will be fewer critical (unpaired) nodes. As the problem of finding optimal discrete gradient vector fields is MAX-SNP hard [Lewiner 02], we proved here that the problem of finding a maximal hyperforest in a hypergraph is also MAX-SNP hard.

4.6 A Topological Invariant for 3-Manifolds

Definition 4.10. (Discrete Morse numbers.) The Morse number $\mathcal{M}_p(K)$ of index p of a cell complex K is the minimum possible number of critical p-cells, considering all possible discrete gradient vector fields defined on K.

Morse theory is linked to simple homotopy. To prove the invariance of the Morse numbers, we could prove that topologically equivalent cell complexes are simple homotopic, and that simple homotopic spaces have the same discrete Morse numbers. Unfortunately, the first affirmation is not true in the general case. However, it holds for 3-manifolds. We will use the following theorems, proofs of which can be found, respectively, in [Moïse 52] and [Cohen 73, 25.1].

Theorem 4.11. (3-Manifold Hauptvermutung.) *Any two triangulations of a topological 3-manifold have a common subdivision.*



(a) Original model (b) C^1 (c) C^2

FIGURE 11. A contractible space and the complement cell complexes C^1 and C^2 of two different hyperforests $H^1_{2/1}$ and $H^2_{2/1}$ defined on it.

Theorem 4.12. *If K_* is a subdivision of K, then K and K_* are simple-homotopy equivalent.*

The proof of the invariance now follows:

Theorem 4.13. (Invariance of discrete Morse numbers.) *Let K and L be homeomorphic 3-manifolds. Then for all p, $\mathcal{M}_p(K) = \mathcal{M}_p(L)$.*

Proof: Let $\mathcal{V}$ be an optimal discrete gradient vector field defined on K. We will prove the theorem by contradiction. Suppose L would have its Morse number of index n greater than the number of critical n-cells of $\mathcal{V}$. We will construct a discrete gradient vector field on L with the same number of critical elements as $\mathcal{V}$.

From Theorem 4.11, there exists a common subdivision to K and L. We deduce from Theorem 4.12 that L can be obtained from K by a finite number of collapses and extensions.

If M_* is an extension of M, and $\mathcal{V}$ is a discrete gradient vector field defined on M, we know from [Forman 98, Section 12] that we can define a discrete gradient vector field $\mathcal{V}_*$ on M_* with the same number of critical elements as $\mathcal{V}$. If M collapses onto M_*, we know from [Forman 98, Lemma 4.3] that we can extend $\mathcal{V}_*$ on M without adding any critical element.

Therefore, we can build a discrete gradient vector field on L with the same number of critical elements as $\mathcal{V}$. This leads to the desired contradiction. □

5. CONSTRUCTING DISCRETE GRADIENT VECTOR FIELDS

The algorithms we will introduce process each layer $\mathcal{L}_{p/q}$ of the Hasse diagram. For each of those, they define a hyperforest $H_{p/q}$ extracted from the layer's hypergraph $\mathcal{H}_{p/q}$, i.e., they define a discrete gradient vector field. The optimizations we perform in the algorithm are local. This is sufficient only for 3-manifolds. Locally, we are able to maximize the number of regular links and loops. Maximizing the number of nonregular hyperlinks is performed by a greedy heuristic.

5.1 Validity of Local Optimization for 3-Manifolds

We proved in 4.6 that the minimal number of critical cells is an invariant at least for 3-manifolds. We will deduce that maximizing the number of hyperlinks in each layer of the Hasse diagram gives a global maximum:

Theorem 5.1. *Let K be a 3-manifold cell complex. Any two optimal discrete gradient vector fields defined on K will have the same number of critical cells in each layer.*

Consider two different $n/(n\text{-}1)$-hyperforests $H^1_{n/(n\text{-}1)}$ and $H^2_{n/(n\text{-}1)}$ extracted from the $n/(n\text{-}1)$ layer of K, having the same number of critical components. Now call C^1 and C^2 the two cell complexes represented by the cells of K of dimensions $\leq n$ and whose $(n\text{-}1)$-cells are not in $H^1_{n/(n\text{-}1)}$ and $H^2_{n/(n\text{-}1)}$, respectively (see Figure 11). From Theorems 3.3 and 3.4 of [Forman 98], C^1 and C^2

are simple homotopic and they have the same minimal number of critical cells. We conclude by induction that, in the case of 3-manifolds, maximizing the number of hyperlinks in each hyperforest generates an optimal discrete gradient vector field.

5.2 Algorithm Outline

We must first choose which layers of the Hasse diagram we process. Actually, we can process all of them, independently from their direct or their dual hypergraph representation. We know that the dual pseudograph $\mathcal{H}_{n/(n\text{-}1)}$ of a manifold has no nonregular hyperlink, and that the direct hypergraph $\mathcal{H}_{0/1}$ of the first layer is a simple graph. Those two simple cases could be useful as the construction of hyperforest is linear on pseudograph and quadratic on general hypergraphs. For example, a solid model could be processed by the following sequence of layers: 0/1,1/2,3/2; or 3/2,2/1,0/1. That is, we extract one hyperforest $H_{p/q}$ of each of the (d-1) successive hypergraph $\mathcal{H}_{p/q}$ (where d is the dimension of the cell complex).

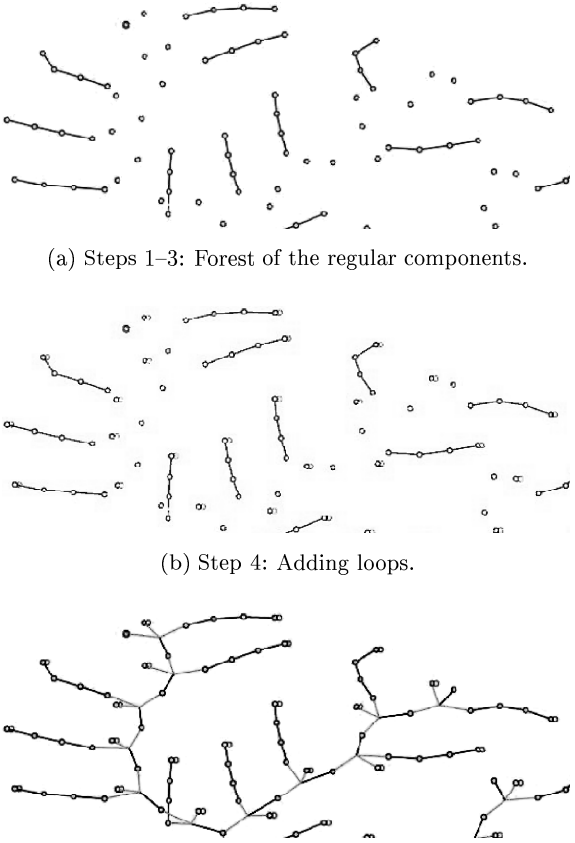
In this work, all the algorithms extract a hyperforest $H_{p/q}$ out of a hypergraph $\mathcal{H}_{p/q}$ process by the following steps (see Figure 12):

1. Initiate $H_{p/q}$ with the nodes of $\mathcal{H}_{p/q}$.
2. Generate a spanning tree on every regular component of $\mathcal{H}_{p/q}$.
3. Add all the links of those spanning trees to $H_{p/q}$.
4. If a regular component is incident to some loops, add one of them to $H_{p/q}$.
5. Add the nonregular hyperlinks of $\mathcal{H}_{p/q}$ which do not create cycles.

The first four steps of the algorithm are linear, and guaranteed to be optimal in any case. The last step requires some heuristics as detailed below.

5.3 Maximum Number of Regular Hyperlinks

Let K be a cell complex, p and q two successive dimensions of K and $\mathcal{H}_{p/q}$ the hypergraph of $\mathcal{L}_{p/q}$, the p/q layer of its Hasse diagram. For any hyperforest $H_{p/q}$ extracted from $\mathcal{H}_{p/q}$, consider RT the simple graph whose nodes are the nb nodes of a regular component R of $\mathcal{H}_{p/q}$, and whose links are the regular hyperlinks of $H_{p/q}$ incident to those nodes. As R is a regular component of $\mathcal{H}_{p/q}$, there is no regular hyperlink incident to a node of R and a node out of R: RT is well defined. As $H_{p/q}$ is a hyperforest, there is no circuit in RT. Therefore, RT is a collection of k trees: RT has $(nb - k)$ links. The maximum number of links will thus be for k minimal, i.e., RT a unique (connected) tree. This optimum can be reached by constructing a spanning tree on each regular component of $\mathcal{H}_{p/q}$ [Hopcroft 73].



(a) Steps 1–3: Forest of the regular components.

(b) Step 4: Adding loops.

(c) Step 5: Adding nonregular hyperlinks.

FIGURE 12. Steps of the algorithm on a part of the hyperforest 2/1 of $S^2 \times S^1$.

5.4 Maximizing the Number of Loops

Each connected component of $H_{p/q}$ is critical or incident to either a loop or a nonregular hyperlink. The problem of the regular hyperlinks has been resolved optimally, and we want now to maximize the number of loops and nonregular hyperlinks of $H_{p/q}$. If a critical component is incident to a loop in $\mathcal{H}_{p/q}$, then adding this loop to $H_{p/q}$ generates another hyperforest with one more hyperlink, i.e., one critical component less. If a regular component is incident to a loop l in $\mathcal{H}_{p/q}$ and to a nonregular hyperlink nl in $\mathcal{H}_{p/q}$ and $H_{p/q}$, then replacing nl by l in $H_{p/q}$ generates another hyperforest with the same number of hyperlinks (and a lower risk to create a hypercircuit).

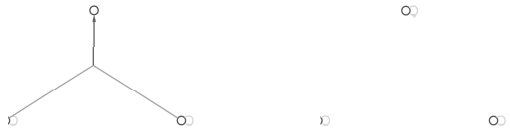


FIGURE 13. Replacing a nonregular hyperlink by a loop.

This process is illustrated in Figure 13. Therefore, we can always generate a hyperforest $H_{p/q}$ with the maximum possible number of hyperlinks such that every regular component incident to a loop in $\mathcal{H}_{p/q}$ is incident to a loop in $H_{p/q}$.

5.5 Condition for Nonregular Hyperlink Insertion

Let $H_{p/q}$ be the hyperforest being created out of the hypergraph $\mathcal{H}_{p/q}$. A hyperlink can be added to $H_{p/q}$ only if it is incident to at least one critical component. Otherwise, $H_{p/q}$ would not be a hyperforest, according to Proposition 4.5. For a nonregular hyperlink nl, let $\mathcal{C}(nl)$ denote the set of connected components of $H_{p/q}$ containing a node incident to nl in $H_{p/q}$.

The hyperlink nl can create a hypercircuit in a connected component C of $\mathcal{C}(nl)$ when it is incident to more than one node of C, or more than one time to a node of C, and when the source of nl is a node of C (see Figure 14).

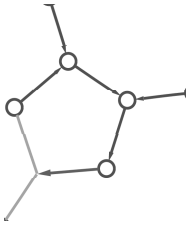


FIGURE 14. A hyperlink creating a hypercircuit.

If there exists a connected component C in $\mathcal{C}(nl)$ such that nl is simply incident to only one node n of C, and if n belongs to a critical component of $H_{p/q}$, then nl can be added to $H_{p/q}$. This node c will be the source of nl. If there does not exist such a node, we can remove nl from $\mathcal{H}_{p/q}$ as it will never belong to $H_{p/q}$. This case is valid if nl is not incident to any critical component.

In particular, when a regular component is incident to only one hyperlink and if the hyperlink is incident only once to this regular component, we can add it to $H_{p/q}$ (see Figure 15).

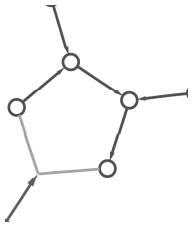


FIGURE 15. A regular component incident to only one hyperlink.

6. FURTHER HEURISTICS

The complexity of the first four steps of our algorithm is linear in time of execution. Each of those steps gives an optimal result. However, the problem of finding an optimal discrete gradient vector field is MAX-SNP hard [Lewiner 02]. Thus, the last step of our algorithm, i.e., deciding which nonregular hyperlinks of $\mathcal{H}_{p/q}$ will be added to $H_{p/q}$, must be much more expensive. If the size of the hypergraph allows it, we could use an exponential algorithm, generating all possible hyperforests and testing which is the maximal one. For the general case, we provide in this section different heuristics together with their results to complete the last step of our algorithm.

6.1 Greedy Methods

Let $H_{p/q}$ be the hyperforest being created out of the hypergraph $\mathcal{H}_{p/q}$. We can try to add the hyperlinks of $\mathcal{H}_{p/q}$ to $H_{p/q}$ in a greedy manner. The criterion for a hyperlink to be added or not to $H_{p/q}$ has been discussed in Section 5.5.

The priority on links can be quite arbitrary, as there is no polynomial approximation. We tested three of them:

- minimal number of incident regular components;
- minimal number of incident critical components in $H_{p/q}$;
- maximal number of incident noncritical components in $H_{p/q}$.

The problem that appears with those criteria is that the priority must be calculated again each time a hyperlink is added to $H_{p/q}$ (as some components change status from critical to noncritical). Therefore, the complexity of such a heuristic is worst-case quadratic.

6.2 Mixing with Geometry

We can impose some more conditions on our discrete gradient vector fields, similarly to [Lewiner 03]. However, there is a difference with that case: The geometry can

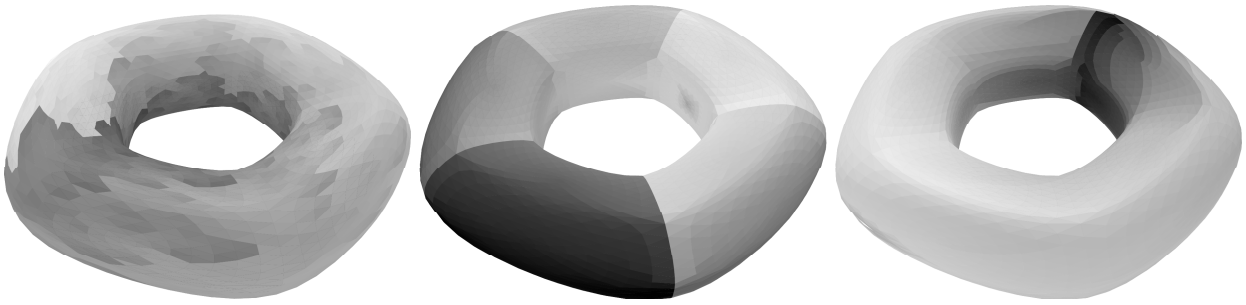


FIGURE 16. Discrete gradient vector fields with geometric constraints.

influence the result, as the hyperforest of a layer will be different if the hyperforest of the previous layer processed is a geometrical minimum.

There are different constraints we can add on our hyperforest $H_{p/q}$:

- The spanning tree of the regular components of $H_{p/q}$ can be chosen to be a minimal spanning tree.
- The loops added to the regular components of $H_{p/q}$ can minimize the same function, in order to have the root of the spanning trees at a minimal position.
- The roots of the spanning trees of the critical components of $H_{p/q}$ can also be at a minimal position.
- The priority used in the greedy heuristics (see Section 6.1) can be derived from the same geometric function.

The discrete gradient vector fields generated on a torus model with those constraints are represented in Figure 16 for different geometric functions. The vector field goes from light gray to dark gray. All of them have four critical points (one vertex, two edges, one face).

6.3 Experiments

We compared our different heuristics on two kinds of models: Hachimori's examples [Hachimori 01] (mainly nonconstructible models), and other solid models at the Mat&Mídia Laboratory (see Table 2). The results of those processes are given in Table 1. The different heuristics we implemented were:

	HG Simpl	Min Deg	Min Def	Max Cpl
Direct	1208	530	14	402
Dual	7258	728	8	934
Sym Direct	3580	658	50	702
Sym Dual	3842	566	6	722

TABLE 1. Number of redundant critical cells per method on the models of Table 2.

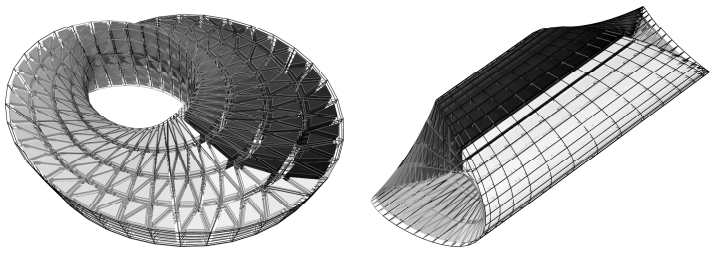
- Direct: processing the layers 0/1, 1/2, 2/3.
- Dual: processing the layers 3/2, 2/1, 1/0.
- Sym Direct: processing the layers 0/1, 1/2, 3/2.
- Sym Dual: processing the layers 3/2, 2/1, 0/1.
- HG Simpl: only simplifying the hypergraph, with no further process.
- Min Def: priority to the hyperlinks incident to the minimum number of critical components.
- Min Deg: priority to the hyperlinks of minimum degree.
- Max Cpl: priority to the hyperlinks incident to the maximum number of noncritical components.

Forcing the first and last layers to be processed as 0/1 and $n/(n\text{-}1)$, as in the cases of Sym Direct and Sym Dual, leads to the best results (see Table 2), because it generates fewer nonregular hyperlinks to be processed: The 0/1 layer is a multigraph (1-skeleton) and, for the case of manifolds, the $n/(n\text{-}1)$ layer is a pseudograph.

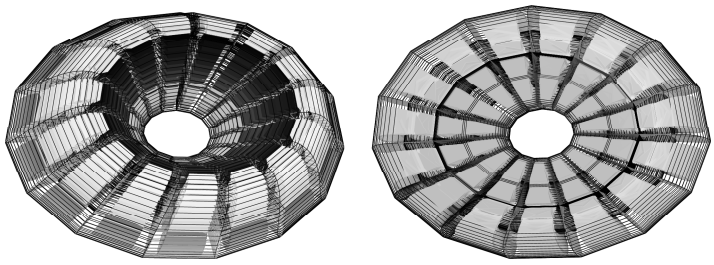
The Sym Dual is usually the best processing order, as it avoids disconnecting the cell complex as would do Sym Direct (for example, a 0/1 spanning tree on a surface with boundary, with two vertices on the same boundary). In particular, for 2-manifolds, the Sym Dual algorithms are optimal [Lewiner 03].

The Min Def priority in the greedy algorithm leads to the best algorithm by far. When looking at the detailed results, mixing with the geometry of the cell complex, when available, considerably improves the performances of the algorithm.

Figures 17 and 18 show the resulting discrete gradient vector field on a four-dimensional model of a Cartesian product of a Möbius strip by a circle. The algorithm used was the Sym Dual/Min Def algorithm, minimizing the total "y" coordinate of the spanning trees and root

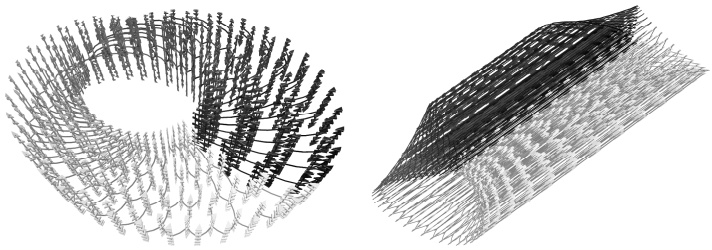


(a) xyz (b) xyw

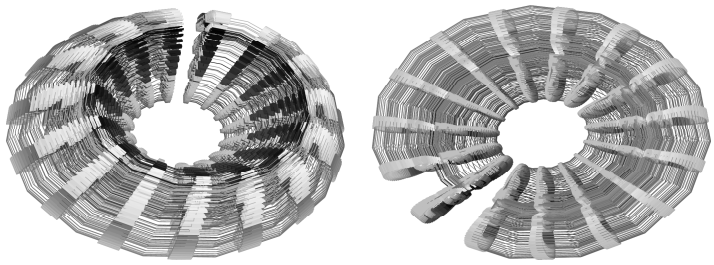


(c) xzw (d) yzw

FIGURE 17. A discrete gradient vector field on a Cartesian product of a Möbius strip by a circle: 2925 0-cells, 10500 1-cells, 12225 2-cells, and 4650 3-cells; 1 critical 0-cell, 2 critical 1-cells, 1 critical 2-cell, and 0 critical 3-cell. The discrete gradient vector field, drawn for the 1- and 3-cells, goes from light gray to dark gray. The 3-cells are shrunken for the sake of clarity.



(a) xyz (b) xyw



(c) xzw (d) yzw

FIGURE 18. The 3/2-spanning tree of the discrete gradient vector field of Figure 17. The 2-cells of the 3/2-spanning tree are drawn as line, and the 3-cells as small solids.

Model	Topology	Number of cells	Euler char	Number of critical cells
bing	3-ball	(480,2511,3586,1554)	1	(1,1,1,0)*
bjorner	projective plane + one facet	(6, 15, 11)	2	(1,0,1)
c-ns	contractible	(12, 37, 26)	1	(1,1,1)
c-ns2	contractible	(13, 39, 27)	1	(1,0,0)
c-ns3	contractible	(10, 31, 22)	1	(1,1,1)
dunce hat	Dunce hat	(8, 24, 17)	1	(1,1,1)
gruenbaum	3-ball	(14, 54, 70, 29)	1	(1,0,0,0)
knot	3-ball	(380,1929,2722,1172)	1	(1,1,1,0)
lockeberg	3-sphere	(12, 60, 96, 48)	0	(1,0,0,1)
mani walkup C	3-sphere	(20, 126, 212, 106)	0	(1,0,0,1)
mani walkup D	3-sphere	(16, 106, 180, 90)	0	(1,0,0,1)
nonextend	contractible	(7, 19, 13)	1	(1,0,0)
poincare	homology sphere	(16, 106, 180, 90)	0	(1,2,2,1)
projective	projective plane	(6, 15, 10)	1	(1,1,1)
rudin	3-ball	(14, 66, 94, 41)	1	(1,0,0,0)
simon	contractible	(7, 20, 14)	1	(1,0,0)
ziegler	3-sphere	(10, 38, 50, 21)	1	(1,0,0,0)
Pile of Cubes	contractible	(572,1477,1266,360)	1	(1,0,0,0)
s2xs1	$S^2 \times S^1$	(192, 588, 612,216)	0	(1,1,1,1)
s3	3-sphere	(162, 522, 576,216)	0	(1,0,0,1)*
solid 2sphere	2-sphere	(64, 144, 108, 26)	2	(1,0,1,0)
Furch knotted ball	3-ball	(600,1580,1350,369)	1	(1,1,1,0)*

TABLE 2. Results on solid models. (* points models for which the best result is not obtained by Min Def/Sym Dual, but by Min Def/Sym Direct).

positions. The discrete gradient vector field goes from light gray to dark gray.

7. OPEN PROBLEMS

This work was focused on Forman's discrete Morse theory. We analyzed the building blocks of this theory, and proved the layered structure of discrete gradient vector field. We represented this layer structure by a collection of hyperforests and gave a complete characterization of the critical cells in terms of regular components of hyperforests. We used this analysis to introduce a scheme for constructing discrete gradient vector fields on finite cell complexes of arbitrary dimension. Although the general problem is MAX-SNP hard, this construction is quadratic in time in the worst case, and is proven to be linear and optimal in the case of 2-manifolds. The experimental results have shown our algorithm gave an optimal result in most of the cases.

We know from the disproof of the Hauptvermutung [Moïse 52] that combinatorial invariants of triangulations are not always topological ones. Thus, the discrete Morse numbers could not be a topological invariant in the general case. However, for the case of 3-manifolds, we proved here that discrete Morse numbers are topological invariants.

Our algorithms seem to be optimal in all the cases we studied, except for the knotted ball and the Bing's house.

Finding the conditions that would guarantee an optimal result in polynomial time remains an open problem.

REFERENCES

[Berge 70] C. Berge. *Graphes et hypergraphes.* Paris: Dunod, 1970.

[Cohen 73] M. M. Cohen. *A Course in Simple Homotopy Theory.* New York: Springer-Verlag, 1973.

[Forman 95] R. Forman. "A Discrete Morse Theory for Cell Complexes." In *Geometry, Topology and Physics for Raoul Bott*, edited by S. T. Yau, Chap. 3. Cambridge, MA: International Press, 1995.

[Forman 98] R. Forman. "Morse Theory for Cell Complexes." *Advances in Mathematics* 134 (1998), 90–145.

[Forman 00] R. Forman. "Applications of Combinatorial Differential Topology." In *Proceedings of the Sullivan-fest.* To appear, 2000. Available from World Wide Web (http://math.rice.edu/~forman/).

[Forman 01] R. Forman. "A User Guide to Discrete Morse Theory." In *Séminaire lotharingien de combinatoire*, 48. Villeurbanne, France: Institut Girard Desargues, 2001.

[Hachimori 01] M. Hachimori. Simplicial Complex Library. Available from World Wide Web (www.qci.jst.go.jp/~hachi/math/library/index_eng.html), 2001.

[Hopcroft 73] J. Hopcroft and R. E. Tarjan. "Efficient Algorithms for Graph Manipulation." *Communications of the Acm* 16 (1973), 372–378.

[Lewiner 02] T. Lewiner. "Constructing Discrete Morse Functions." Master's thesis, PUC-Rio, 2002. Available from World Wide Web (http://www.mat.puc-rio.br/~tomlew/Thesis.pdf).

[Lewiner 03] T. Lewiner, H. Lopes, and G. Tavares. "Optimal Discrete Morse Functions for 2-Manifolds." *Computational Geometry: Theory and Applications* 26:3 (2003), 221–233.

[Lovász 86] L. Lovász and M. D. Plummer. *Matching Theory.* Amsterdam: Van Nostrand Reinhold, 1986.

[Lundell 69] A. Lundell and S. Weingram. *The Topology of CW Complexes.* New York: Van Nostrand Reinhold, 1969.

[Milnor 63] J. W. Milnor. *Morse Theory.* Princeton, NJ: Princeton University Press, 1963.

[Moïse 52] E. E. Moïse. "Affine Structures in 3-Manifolds." *Annals of Math* 56:2 (1952), 96–114.

Thomas Lewiner, Laboratório MatMídia, Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil (thomas.lewiner@polytechnique.org) and Project Géométrica, INRIA—Sophia Antipolis, France.

Hélio Lopes, Laboratório MatMídia, Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil (lopes@mat.puc-rio.br)

Geovan Tavares, Laboratório MatMídia, Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil (tavares@mat.puc-rio.br)

Received August 6, 2002; accepted in revised form July 24, 2003